DOCUMENT RESUME

ED 195 095 EC 130 893

AUTHOR Hanna, Gerald S.: And Others
TITLE Comparing Potential with Achievement: Rationale and Procedures for Objectively Analyzing Achievement-Aptitude Discrepancies in LD Classification.
INSTITUTION Kansas State Univ., Manhattan.
PUB DATE [80]
NOTE 39p.: Portions of this manuscript contain material abridged in an article in the "Learning Disabilities Quarterly (Fall, 1979)" as well as tables that render the previously described procedures operative.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Aptitude Tests: Elementary Secondary Education: *Evaluation Methods: *Handicap Identification: *Learning Disabilities: Literature Reviews: *Reading Achievement: *Reading Difficulties: Test Interpretation

ABSTRACT

A critical review of the literature dealing with quantification of achievement-aptitude differences for identifying learning disabled (LD) readers revealed that methods developed to date suffer from grave inadequacies. Among the methods considered were those of the following individuals: G. Bond and M. Tinker, M. Monroe, A. Harris and E. Sipay, H. Myklebust, C. Winkley, C. Woodbury, and M. Erickson. Based upon the evaluation of past methods, criteria for new and more adequate procedures were generated and discussed. Ten requirements for quantifying achievement-aptitude descrepancies were specified: (1) all derived scores should be based on national, rather than local, reference groups: (2) national normative data from well standardized aptitude and achievement tests may be accepted as roughly comparable: (3) the relevance of an aptitude test to the subject field for which it is used to compare aptitude and achievement should depend on the kind of educational intervention being considered: (4) comparable kinds of derived scores must be used to report performance on achievement and aptitude tests and these derived scores must be based on interval scales: (5) the kind of scores used should be available or conveniently derivable for the normative data available for commonly used standardized tests: (6) age based and grade based normative data should not be mixed: (7) all scores used must avoid the illusion of ratio scaled: (8) the standard error of the difference between grade based T-scores for achievement and aptitude should be found: (9) there should be no rigid cutoff applied concerning the amount of achievement-aptitude discrepancy necessary for LD classification: and (10) in classifying LD students, errors of commission may be less serious than errors of omission. Step by step procedures, with examples, for quantifying achievement-aptitude discrepancies are set forth. (Author/SBH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

COMPARING POTENTIAL WITH ACHIEVEMENT: RATIONALE AND PROCEDURES FOR OBJECTIVELY ANALYZING ACHIEVEMENT-APTITUDE DISCREPANCIES IN LD CLASSIFICATION*

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Gerald S. Hanna
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Gerald S. Hanna Norma J. Dyck Michael C. Holen
Kansas State University

ED195095

Faced with the necessity of judging which students might benefit most from special remedial or compensatory instruction, special educators are confronted with the practical need to compare students' academic aptitude and achievement. But as knowledgeable practitioners have long recognized, attempts to make this seemingly simple comparison lead one through a veritable minefield of measurement hazards. Writers have proposed a perplexing variety of methods of comparing aptitude with achievement. Unfortunately, these methods contain a myriad of major problems, serious errors, and gross inpracticalities.

Yet there is a continuing need for a viable objective means of assessing achievement-aptitude discrepancy. The urgency of this need is evidenced by the regulations for PL 94-142. The law mandates that a significant achievement-aptitude discrepancy must be observed for placement in a LD program, but procedures for determining the discrepancy are not provided.

This article is designed to enable sounder comparisons of students' current academic potential with their achievement. In pursuit of this objective, the authors have (1) critically reviewed past methods (to locate some of the mines in the field), (2) formulated the rationale for a sounder method of quantifying achievement-aptitude discrepancies, (3) devised specific

EC130893

*Portions of this manuscript contains material abridged in an article in the Learning Disabilities Quarterly, Fall 1979, as well as tables that render the previously described prodedures operative.

step-by-step procedures by which to implement their proposals (thereby avoiding the more dangerous mines), and (4) provided tables that are needed to make the necessary conversions.

Review of Literature

This review consists of an evaluation of several quantitative procedures that have been used to compare academic potential with achievement. Although the focus of most of the literature has been upon reading, these methods are easily generalized to other school subjects. The procedures described and evaluated were selected for inclusion because of their popularity and/or their variety; problems attending other methods with which the authors are familiar are highly similar to those treated in this review. The purpose of this selective, critical literature review was to lay the foundation for better methods of assessing aptitude-achievement discrepancies.

Classical Discrepancy Method

One of the simplest methods by which to identify disabled readers is to add 5 years to the student's grade equivalent (GE) to obtain an estimated age equivalent (AE) (e.g., Harris, 1961; Winkley, 1962). The discrepancy between a student's expectancy and achievement is then obtained by subtracting the achievement AE from the mental age (MA). This difference must then be interpreted in light of the individual's grade level placement because a given discrepancy (e.g., .8) would be more substantial and important to a pupil in the lower grades than to a student in the upper grades.

This method suffers from several intolerable problems. First, the use of MAs is unjustifiable in view of their widely recognized limitations. Among these problems are (1) MAs cannot legitimately be subjected to arithmetic computations, such as the subtraction used in this method, because they lack a consistent unit of measurement, (2) since MAs are meaningless beyond the years

of rapid, relatively linear intellectual growth, methods based upon their use are not applicable at the high school or adult level, and (3) MAs do not focus upon the comparison of examinees with their most appropriate reference group--their grade or age peers. Acknowledgement of the severe problems associated with the use of MAs has led to their abandonment, for purposes of computing IQs, in virtually all major tests published or revised since about 1960. Deviation IQs(DIQs), which do not use MAs, have largely replaced the obsolete ratio IQs.

Second, GEs have similar widely publicized and recognized measurement features that render them wholly inappropriate for use in comparing scores among tests. Among these features are their (1) provision of normative data by which examinees are compared with groups to which they do not belong, (2) inconsistent variability across different kinds of content, e.g., arithmetic and reading, with the concominant hazards in profile interpretation, (3) uselessness beyond the growing years, and (4) unequal units of measurement with their consequent illigitimacy in arithmetic calculations, e.g., computation of means, differences, or quotients. These and other limitations of GEs are detailed in most introductory texts in educational measurement. Along this line, Cronbach, (1970, p. 98) concluded that GEs "should never be used in reporting on a pupil."

Third, the simple discrepancy method seems to be based upon the unstated assumptions that all students enter school at the same age and progress through the grades at the same rate. Clearly these assumptions are untrue. Yet the use of a formula that mixes and compares age-based MAs with grade-based GEs is predicated upon these untenable assumptions.

Fourth, the method seriously overestimates the potential of young, high IQ children. No account is taken of the fact that specific school

learning is a function of amount of time in school as well as a funciton of aptitude. For example, it is not realistic to expect a child 6-0 years old with an IQ of 140 to achieve a GE of 3.4 upon entrance to first grade just because his or her "mental age" is about 8-5.

Fifth, different interpretations must be made for a given discrepancy depending upon the GE level at which it occurs (not merely upon the student's grade placement level as some writers have indicated). This nuisance arises from the fact that GEs and MAs are only ordinal measurement scales.

The method has but one attractive feature. It is simple. Persons unencumbered by any knowledge about educational measurement may quite easily and very erroneously think they understand the method. This is not a virtue!

Bond and Tinker's Method

Bond and Tinker (e.g., 1973) used the expression $\frac{IQ}{100} \times$ number of years in school + 1.0 to obtain a measure of reading potential expressed in GEs. In computing years in school, repeated grades are counted double while kindergarten is ignored. This estimate of potential is then compared with the student's GE.

Problems with this method involve all but the fourth of those enumerated for the classical discrepancy method. Namely, IQs are treated as though they were the obsolete ratio variety and as though all aptitude tests yield scores having the same standard deviation, GEs are used, age-based and grade-based data are mixed, and different criteria of deficiency are needed at different levels.

The main difference between the methods concerns the computation of growth. The Bond-Tinker method begins with the point at which the pupil starts first grade. While the classical method erred in assuming that specific school learning is independent of amount of time spent in school, the Bond-Tinker

method is based upon the opposite assumption that school learning is not a function of amount of time lived. Neither of these points of view seems attractive. The truth may be that one's level of specific learning reflects both amount of time lived and time spent receiving formal instruction.

Although the Bond-Tinker procedures differ from the classical discrepancy method in this relatively minor dimension, they retain all of the condemning features of the older method.

Monroe's Index

Monroe (1932) estimated reading expectancy by use of the mean of the student's chronological age converted into a GE, mental grade, and arithmetic GE. This average was then divided into the examinee's reading GE to obtain the Reading Index.

This method, like Bond and Tinker's, suffers from the first three problems of the classical discrepancy method. These faults render the method wholly unsuitable. In addition, the unfortunate use of a quotient gave rise to the fallacy of treating the index values as though they constituted ratio measurement scales, i.e., as though there were a meaningful absolute zero point as there is in measures of such variables as time, money, and weight. For example, "Mary's reading is . . . only 56 per cent of the expectation . . ." (Monroe, 1932, p. 14). Dividing values of one ordinal scale by values of another ordinal scale will not produce a ratio scale!

Given the point in time when this method was devised, its avoidance of the fourth and fifth problems of the classical discrepancy method seems laudable. Moreover, the unacceptable features of GEs and MAs were not widely recognized in the 1930's.

Harris and Sipay's Method

Harris and Sipay (1975) recommended that Reading Expectancy Age (R Exp. A) be obtained by the expression $\frac{2MA+CA}{3}$. The student's Reading Expectancy Quotient then equals $\frac{RAX100}{R\ Exp.\ A}$ where RA is the reading AE. Quotients below 90 are said to indicate disability.

The problems involved with the use of this method include multiple inappropriate uses of MAs and achievement ages and the masquerading of ordinal scales as ratio scales. In addition, it is unwise to focus selection of disabled students upon a fixed numeric value, such as Harris and Sipay's 90. In view of both the imperfect reliability of the tests used and of the need to consider additional sources of information in making classification decisions, it is naively simplistic to set and use rigid cutting points.

This method and several similar methods not included in this review have very little to offer over the procedures developed by Monroe over forty years earlier. Such anachronistic procedures fail to reflect and use the widely disseminated advances in educational measurement of the past generation.

Myklebust's Learning Quotient Method

One final approach will be considered from those that compare AEs and GEs as though they all had identical variability and use ordinal scales as though they were interval or ratio scales. Myklebust (1968) suggested that the expectancy age should be found by computing the mean of MA, Life Age, and Grade Age. In this way something of a compromise is achieved between the neglect of the classical discrepancy method to allow for time in school and the neglect of the Bond-Tinker method to allow for learning prior to the onset of formal schooling.

Another suggestion by Myklebust was that both verbal and nonverbal tests be used to assess aptitude. This notion could have led to the discriminating use with each student of the aptitude test that is deemed most relevant to the specific learning disability being considered. But Myklebust instead directed the recommendation toward routinely using the higher of the two scores. This provides two opportunities for chance errors of measurement to cause aptitude to be overestimated. Use of the highest of several kinds of aptitude measures is at times justified in identifying learning disabled students for compensatory interventions, but has little to commend it for identifying LD students for remediation interventions.

Myklebust also recommended that discrepancy between achievement and aptitude be determined in each achievement area of concern, not just in reading. This commendable recommendation is widely accepted by authorities in the field of specific learning disabilities.

In other respects, Myklebust's approach is as unsophisticated as those reviewed above. The student's learning quotient is computed by dividing the achievement age by the expectancy age; "90 percent of average should be the baseline in computing expectancy for learning" (Myklebust, 1968, p. 6).

Problems with this approach involve use of AEs, lack of consideration of differing standard deviations in AEs among tests of different functions, use of the ratio of achievement to expectancy which is very similar to the achievement quotient that had long since fallen into disrepute and disuse because of widely recognized logical and statistical shortcomings (Thorndike, 1963, p. 45), the assumption that the quotient of values from two ordinal scales magically yields a ratio scale value, and the use of an arbritrary cut-off point.

Winkley's Stanine Method

Winkley (1962) computed, for research purposes, achievement and aptitude stanines from local data. She then used a difference of two or more stanines between achievement and potential as an indicant of reading disability.

One problem concerns the use of the arbitrary two-stanine criterion. Because of the half-standard-deviation width of stanines, the requirement of a two stanine difference can be met in some cases with as little as a .51 standard deviation difference while in other cases a difference as large as a 1.00 standard deviation would not meet this criterion. This needless grouping error is too great to tolerate.

The use of local groups for computing normative data forfeits the broad base of reference-group data available for nationally standardized tests. To illustrate this point with an extreme example, suppose that a local school system did not start instruction in reading until fourth grade. Winkley's local-norm method would not reveal the fact that a large fraction of this district's fourth-grade pupils were far below expectancy; but a system based upon national reference-group data would dramatically reveal this situation.

On the positive side, Winkley's approach was among the first to avoid the unacceptable uses of MAs and GEs. In this respect, the method constitutes a major improvement over all previously considered approaches.

Woodbury's Method

Woodbury (1963) used, for research purposes, a method whereby (1) equipercentile scaling of achievement and aptitude scores is performed on a local population of students, (2) the aptitude scores are regressed toward their mean by use of the regression coefficient based on the correlation of achievement and aptitude scores, (3) each student's regressed aptitude score is subtracted from his or her achievement score, and (4) the standard error of estimate is used as the criterion of necessary difference for classification.

The use of local populations to render achievement and aptitude test scores comparable is simultaneously a virtue and a fault. On the one hand, it provides an escape from the need to use scores based on different standardization samples; the local group becomes the reference group for both tests. On the other hand, many local populations, including Woodbury's 134 cases, are far too small to provide dependable comparison among scores of various tests.

The most serious fault of Woodbury's method concerns the use of a regression equation. This is an appropriate technique by which to estimate (or predict) the scores of one test from the scores of another test. But in considering specific learning disabilities or underachievement, both scores are known; it is necessary only to compare them. Regression equations and standard errors of estimate are suitable for the prediction of future unknowns. Comparisons of comparable scores and standard errors of measurement of differences are appropriate in the profile analysis of presently available knowns.

Finally, Woodbury's procedures are torturously laborous. This would discourage their use in most applied settings.

Perhaps the greatest advance of this method was its avoidance of inappropriate uses of age and grade equivalents. These procedures achieved some of the benefits that can be secured by use of standard scores and thereby helped to pave the way for their use.

Erickson's Method

Erickson (1975) proposed a standard score discrepancy method of identifying reading disabled students. First, IQ and achievement test scores are secured and the mean and standard deviation of each for the local students is computed. Next, each examinee's scores are converted to standard scores. Finally, each student's standard score for the IQ test is subtracted from his or her standard score for the achievement test.

The only serious criticism these procedures warrant concerns the use of the local population to derive the standard scores. As noted in connection with Woodbury's method, this provision, to its credit, circumvents the problems of the two tests not being standardized on the same sample. However, this advantage is secured at a cost in (1) labor and complexity of computing standard scores for each examinee that would cause all but the most robust of practioneers to flee from its use and (2) sacrifice of the broader normative data available from national standardizations of the tests.

Erickson's method also suffers from the problem of using a kind of conversion, z-scores, that requires the use of negative numbers and decimal fractions. Negative z-scores are especially prone to cause computational errors in the subtraction of aptitude scores from achievement scores.

In addition, this and the two preceding methods could not be used for tests that are unsuitable or impractical for all students. The methods make no provision for use of specialized tests on an as-needed basis.

Erickson's method represents another step in the evolution from Winkley's and Woodbury's earlier efforts. The newer procedures achieved the benefits of the older methods and at the same time (1) reduced the labor of rendering achievement and aptitude scores comparable by use of linear rather than equipercentile scaling, (2) avoided the labor and the error of regressing one set of scores, and (3) avoided the rigid use of a standard error as a minimum difference required for classification. Erickson's procedures further helped to point out the way to better methods of securing quantative data for classification decisions.

Rationale for Quantifying Achievement-Aptitude Discrepancies

The above critique serves not only to reveal the acute need for a sound general method, but more importantly, to reveal assets and liabilities of past

methods in order to show the way to new and better procedures. In this section, ten requirements for a sound general method of quantifying achievement-aptitude discrepancies will be specified and the rationale for each requirement will be discussed.

1. All derived scores used should be based on national, rather than local, reference groups. There are three reasons for this specification. First, it is desirable to be able to compare a student's achievement and aptitude with respect to large and representative reference groups. If some pecularity of local curriculum, instruction, or pupulation were to result in large numbers of youngsters achieving in a subject above or below expectancy, then it would be desirable for quantative methods to reveal this anomaly. But if only the local reference group were used, then local pecularities would be obscured rather than revealed.

Second, tests of nonequivalent content would require separate scaling for each grade. Yet many local school districts lack sufficient enrollment to enable the scaling of achievement and aptitude scores for each grade.

Third, even where enrollment is sufficient to enable local normative data to be collected and scaled for each grade, it would be excruciatingly inconvenient to standardize locally all specialized tests, many of which require individual administration, for each grade. It would require enormous amounts of student time to restandardize every low-frequency specialized test in every district in which it might be needed for one or more students. Finally, standard score equivalents of each raw score on each test would have to be computed locally; this requirement alone would deter most potential users.

2. National normative data from well standardized aptitude and achievement tests may be accepted as roughly comparable. This is a bitter pill to

swallow. The authors have each long warned students of the well known hazards of treating scores from different normative groups as though they were necessarily comparable. But one has no viable alternative. Unless comparability among well standardized instruments is assumed, then such tests as the Wechsler scales and most achievement measures cannot be used in assessing achievement-aptitude discrepancies. Only a few achievement tests (e.g., Sequential Tests of Educational Progress and Iowa Tests of Basic Skills) were simultaneously standardized with aptitude tests. For this reason none of the methods of comparing achievement and aptitude reviewed above were faulted for mixing data from two standardization samples; doing so is unavoidable unless one either restricts oneself to the few tests that were standardized on the same national sample or uses local samples for securing normative data. Since neither of these alternatives is practical, one must either assume comparability or abandon the assessment of achievement-aptitude discrepancy.

Fortunately, the pill may prove less harmful than bitter. One hopes that sampling errors among carefully normed tests would cause no more than minor distortions of achievement-aptitude comparisons.

This consideration places greater importance on the issue of adequacy of normative data in evaluating tests for possible use in estimating achievement-aptitude discrepancies than for other uses of the same tests. Persons who assume comparability of scores should carefully evaluate the representativeness and comparability of each instrument's standardization sample.

3. The relevance of an aptitude test to the subject field for which it is used to compare aptitude and achievement should depend upon the kind of educational intervention being considered. Strategies of intervention for LD students are most often designed to remediate the disability. On the

other hand, there are cases for which plans of compensation for unremediable disabilities are more appropriate.

"Shotgun" procedures are to be avoided. In such an approach, a battery of aptitude tests is given and the one on which the student scores highest is automatically used in comparing achievement and aptitude. This indiscriminant use of any aptitude measure ignores the purpose of using aptitude tests--to predict the degree to which a disability is remediable and/or compensatable.

3A. In the more common case where a student is being considered for LD classification with prospects of remediation, then an aptitude measure should be as relevant as feasible to the subject field for which it is used to estimate achievement-aptitude discrepancy. In the past, IQs have been used to estimate students' expected level of achievement with the apparent assumption that a global IQ was the best available indicant of expected rate of progress for any and all subject fields. We contend that aptitude tests that are most relevant to one subject may not be equally appropriate for determining achievement-aptitude discrepancies in other subjects. For example, a verbal aptitude test would be more appropriate than a nonverbal one in judging the prognosis for remediation of a reading comprehension achievement-aptitude discrepancy. On the other hand, a quantative test would be preferable for estimating potential remediation in arithmetic. Although general omnibus academic aptitude tests may be useful for computing some achievement-aptitude discrepancies, when an aptitude test is available with content more relevent to judging the student's potential for remediation of a possible specific disability, then the more relevant test should be used.

3B. In the less common case where a person with a poor prognosis for remediation is being considered for LD classification, then an aptitude test should be selected for the person's strongest aptitude that could be used in

a compensatory strategy. Regardless of whether remediation or compensation is being considered, one needs an aptitude test that is maximally relevant to what one wishes to predict. When contemplating remediation, an aptitude test useful in predicting success in the deficient subject is needed. When considering prognosis for a compensatory strategy, relevance to the anticipated means of compensation becomes paramont.

4. Comparable kinds of derived scores must be used to report performance on achievement and aptitude tests and these derived scores must be based on interval scales. A sound system of comparing achievement and aptitude cannot be based on kinds of scores that lack equal units of measurement, e.g., MAs, ratio IQs, GEs, or percentile ranks. This requirement tends to narrow the practical choice to various kinds of standard scores.

5. The kind of scores used should be available or conveniently dervi-able for the normative data available for commonly used standardized tests. A method must not only have sound psychometric and statistical properties but it should also be applicable to most tests used in LD classifications. This specification eliminates nonnormalized standard scores because they are uncommon in achievement tests and are inconvenient to compute. Since percentile ranks accompany most standardized tests (and the publishers of the few tests that lack them could ordinarily supply them to users rather easily on a request basis), a single table can be used to convert percentile ranks of any test into normalized standard scores. Such a table is provided in Appendix A.

This criterion narrows the choice to (1) normal curve equivalents, (2) stanines, (3) new or exotic types of normalized standard scores, or (4) T-scores. Normal curve equivalents have the disadvantages of having a very peculiar size of standard deviation, 21.06, and of having some numeric values

that are deceptively similar to those of percentile ranks. Because stanines have only nine levels, each interval is relatively large and the resulting grouping error is substantial; in a situation calling for comparison of scores, reliability is a scarce commodity of paramount importance that one can ill afford to squander with unnecessarily large grouping errors. The world does not need new or exotic varieties of standard scores; there is an overabundance already! Finally, T-scores are not afflicted with any of the above disadvantages or from others (e.g., the need to use fractions or negative numbers) that would hinder their effectiveness. T-scores are therefore the derived score of choice and are supplied in Appendix A.

6. Age-based and grade-based normative data should not be mixed. Aptitude test scores, such as deviation IQs, ordinarily provide comparisons only with examinees' age peers. This tradition is well founded in both theory and practice and is unlikely to change. Achievement test scores, in contrast, typically yield comparisons with the examinees' grade peers (or comparisons with average students in various grades in the case of GEs). These methods of reporting normative information for achievement tests are well established; it is neither desirable nor likely that achievement tests commonly be accompanied with age-referenced norms. As long as one is using only one of these kinds of scores at a time, the inconsistency creates no problems. But it is not appropriate to compare age-based normative data with grade-based data.

Persons needing to compare aptitude and achievement test scores are burdened with this mismatch of reference groups. Although this mismatch tends to have only minor consequences with students whose age is approximately average for their grade, it is a serious problem for students who are markedly younger or older than their grade peers. One must either engage in this kind of illegitimate comparison which falsely assumes all students to be "at age" for

their grades or one must obtain or estimate the same kind of scores for both kinds of tests. Clearly, the latter alternative is preferable.

Since schools normally organize students into grade groups rather than age groups per se, it seems best to opt for grade-based referencing of both achievement and aptitude test scores. For this reason, even persons working in ungraded schools have little choice but to use grade-based scores. In such cases students' "grade" (for purposes of comparing aptitude and achievement) can be based upon their number of years in school.

Alas, the use of grade-based referencing of achievement and aptitude tests is more easily said than accomplished. Once one accepts the goal of using a national sample of examinees in the appropriate part (e.g., fall or spring) of the student's grade as the reference group with which to compare his or her achievement and aptitude test T-scores, one must secure the needed T-scores or percentile ranks that can be converted into T-scores.

For achievement tests, this task is ordinarily very easy. The manuals of well prepared achievement tests virtually always contain grade-based percentile ranks. For the very few achievement test manuals that (deplorably) supply no kind of derived scores except GEs, the user can ask the publisher for grade-based percentile ranks or T-scores. This consumer demand is modest and reasonable. Unless some such type of grade-peer-based derived score is made available, the test cannot be competently used for the assessment of achievement-aptitude discrepancy.

For aptitude tests, securing grade-peer-based derived scores (for purposes of comparing aptitude and achievement) may be more difficult. The use of age-based data for aptitude tests is sound for many purposes. The needs arising for the purpose of studying discrepancy between achievement and aptitude may be the exception rather than the rule. Under these circumstances,

consumers cannot as reasonably demand that publishers satisfy their needs. But they certainly can still ask! Happily, some aptitude tests (e.g., Otis-Lennon Mental Ability Test and Cognitive Abilities Tests) supply grade-based percentile ranks. In using such tests, all that one need do is use Appendix A to convert these grade-based percentile ranks into corresponding grade-based T-scores.

Unhappily, no such easy solution exists for most individual intelligence tests. For example, the Wechsler Intelligence Scale for Children--Revised Edition and the Stanford-Binet Intelligence Scale provide only age-based DIQs. When using such a test, one should seek from the publisher grade-based percentile ranks. If this request is met, then one need only use Appendix A to convert these percentile ranks into T-scores. But if the publisher does not make grade-referenced scores available, then one must either resort to estimating them or abandon use of the test.

The best that one can do to estimate grade-based scores is to find the national median chronologcial age for students in the grade and time of year the aptitude test was administered. Appendix B supplies this information. A student's estimated grade-based aptitude score (for use in determining discrepancy only) can then be found in the age-based table that would be appropriate for a median-aged student in the examinee's grade and time of year. If the score is a percentile rank, Appendix A is then used to convert it into a T-score. If the score is a DIQ, Appendix C can be used to obtain the corresponding T-score. Appendix C will give exact conversions of DIQs into T-scores for tests having normalized DIQs. However, for tests having nonnormalized DIQs, e.g., Wechsler Scales, Appendix C can be used only to <u>estimate</u> T-scores. For such tests, one should first seek from the publisher age-based percentile ranks for use with Appendices A and B. Only if the

publisher also declines to provide these should one resort to the use of Appendix C to estimate T-score equivalents of nonnormalized DIQs.

These procedures are based on the implicit assumption that the results of using median-age-for-grade normative tables will be identical to what would result from the use of grade-based tables if the latter were available. Although this assumption is untrue, it is nevertheless far superior to the alternative and prevailing implicit assumption that age-based and grade-based reference groups yield identical results for all students.

7. All scores used must avoid the illusion of ratio scales. A difference or quotient between two measures can be of no higher measurement scale than the lower of the original measures. Thus ratio IQs are only ordinal scales (although they falsely imply ratio level to naive users), the differences or quotients between CEs and AEs (in addition to being based on mismatched reference groups) are only ordinal scales, and the differences or quotients between standard scores are only interval scales. In order to avoid the implication of ratio scales (as evidenced by such incorrect statements as "Helen is only achieving in arithmetic at 85% of her potential"), quotients should be avoided. Differences are equally useful, have less severe measurement problems, and are less apt to be misunderstood.

8. The standard error of the difference between grade-based T-scores for achievement and aptitude should be found. Although neither this nor any other arbritrary figure should be blindly applied as a cutoff provision for LD classification, one would be remiss in not having a fairly accurate idea of what the likelihood is that a given achievement-aptitude discrepancy between obtained scores was the result of chance rather than a difference between true scores. Appendix D provides a convenient means of obtaining

the approximate standard error of the difference, expressed in T-score units, when a reliability coefficent of each measure is known.

Whether or not a student's achievement-aptitude difference is greater than, say, 1.0 standard error of the difference will be a joint result of imperfect reliability of the two tests and of the true discrepancy between the student's achievement and aptitude. It is therefore strongly recommended that the standard error not be rigidly used as an arbritrary cutting point for LD classification. If only rough screening tests have been used, then less difference than 1.0 standard error of the difference could very well merit further study. On the other hand if unusually reliable tests have been used, then a difference slightly greater than 1.0 standard error may not be of enough practical imprortance to merit classification.

9. There should be no rigid cutoff applied concerning the amount of achievement-aptitude discrepancy necessary for LD classification. There is no empirical or rational answer to the question: "What percent of the nation's students are disabled in reading (or arithmetic, etc.)?" The question is as meaningless as the one asking the percent of people who are underweight or rich. Underweightness or wealth, like LD, is a matter of degree and a matter of definition. If only 1/2% of the population of fifth-grade students is classified or defined as underweight, then "underweightness" signifies marked deviency. On the other hand if 20% of the population is defined as being underweight, then "underweightness" may mean nothing more than slimness. Since there is no one best percentage definition of LD, underweightness, or wealthiness, it may make the most sense simply to work within the prevailing national guidelines that enable a school district to receive compensation for a maximum of approximately 2% of its students.

10. In classifying LD students, errors of commission may be less serious than errors of omission. No great harm is done to the student who is erroneously classified and treated for a disability that turns out to be non-existent. In fact a competent LD teacher will normally discover the error of classification fairly early and cause it to be corrected. On the other hand, considerable harm is done to the truly disabled student who is erroneously considered not to be disabled; the LD may be a lifelong hinderance.

Of course it would be ideal to avoid classification errors altogether. But since tests and other sources of information lack perfect reliability and validity, errors will always occur; the only choice one has is whether to make more of one kind of error or more of the other. It is here argued that it is wiser to err more often in incorrectly assigning students for LD instruction than in incorrectly denying them the benefits of such resources.

This viewpoint leads to a willingness to treat an achievement-aptitude discrepancy as worthy of action in some cases where the difference is not great enough to yield a given level (e.g., .05) of statistical confidence.

Procedures for Analyzing Test Scores in LD Classification

The primary criterion for LD identification concerns the presence of a marked difference between a student's estimated achievement and estimated aptitude or potential wherein the main cause of this discrepancy is not another handicapping condition. In this section, methods of comparing aptitude and achievement are presented as a series of mechanical steps that can be routinely followed. A work-sheet form on which to report and compare test scores is presented and its use in summarizing and analyzing data secured by following the procedures is illustrated.

A secondary criterion for LD classification that is considered by many professionals concerns intraindividual differences in achievement among

various subjects such as listening comprehension, arithmetic computation, spelling, and oral expression. It is widely believed that a student who, in addition to meeting the primary criterion of having higher estimated aptitude than achievement, has higher achievement in one or more other subjects may have a better prognosis for benefiting from LD programs. Since this issue of achievement profile analysis is not a primary criterion for LD classification and since it is not universally used, this criterion is much less important than that of achievement-aptitude differences. Because of its lesser importance and its lesser complexity, its rationale was not developed above.

The step-by-step procedures for analyzing achievement-aptitude discrepancies will first be set out and then illustrated. Next, the rationale and procedures for considering differences among achievement tests of several different subject areas will be presented. Finally, the joint use of these two criteria for LD classification will be discussed and illustrated.

Steps in Quantifying Achievement-Aptitude Discrepancies

I. If a student is suspected of having a learning disability, then measure his or her (1) achievement in the subject under consideration and (2) relevant aptitude.

   A. If the educational intervention for the LD would be remedial in nature, then the aptitude test should be as relevant as practicable to the subject matter in which remediation would be sought. E.g., a verbal aptitude test would be germane to remedial reading.

   B. If a compensatory strategy is being considered, then the aptitude test should be as relevant as feasible to the means of compensation.

II. Report the student's achievement test score in T-score units compared with a national sample of other students in the same grade and time of year.

A. If the achievement test manual provides grade-based T-scores, use them.

B. If T-scores are not available, then obtain the student's grade-based percentile rank and use Appendix A to convert it into a grade-based T-score.

III. Report the student's relevant aptitude test score in T-score units compared with other students in the same grade and time of year.

A. If the aptitude test is accompanied by grade-based T-scores, use them.

B. If grade-based T-scores are not available, then seek grade-based percentile ranks. If available, obtain the student's grade-based percentile rank and use Appendix A to convert it into a T-score.

C. If no grade-based scores are available, then (for purposes of making achievement-aptitude comparisons only):

1. Use Appendix B to find the median chronological age of the country's students for the grade and time of year in which the aptitude test was administered.

2. Estimate the student's grade-based aptitude test score by using the age-based table in the test manual that is suitable for a median-aged student in the examinee's grade. Then:

a. If this grade-based score is a percentile rank, use Appendix A to convert it into a T-score.

b. If this grade-based score is a deviation IQ, use Appendix C to find or estimate the corresponding T-score.

IV. Subtract the grade-based T-score of the aptitude test from the grade-based T-score of the achievement test.

V. If the difference is negative (i.e., if the achievement test T-score is less than the aptitude test T-score), then jointly judge the extent of its significance and the magnitude of the difference:

A. Consider the size of the difference to judge the extent to which the criterion of a marked discrepancy is met

   1. If the difference is less pronounced than about -4 T-score points, it is probably not large enough to satisfy the criterion.
   2. If the difference is between approximately -4 and -8, then it is indecisive whether the criterion is met.
   3. If the difference is more pronounced than about -8 points, it is probably great enough to meet the criterion.

B. Consider the statistical significance of this difference

   1. If the difference, ignoring its algebraic sign, is less than approximately .5 to 1.0 standard errors of measurement for the difference, then evidence of a true difference is relatively weak.
   2. If the difference is between about 1.0 and 1.5 standard errors of the difference, evidence of a true difference is moderate.
   3. If the difference is greater than about 1.5 to 2.0 standard errors, evidence of a true difference is relatively strong.

The results of this comparison of achievement and relevant aptitude test scores is only one of the factors to be considered in LD classifications. The extent to which a student meets this criterion of LD classification should be considered in combination with other factors.

An Example

Before discussing the next criterion, consider the example of Harriet. When she was in the last month of the third grade (i.e., grade 3.9), she received a grade-based percentile rank of 16 on the Reading Comprehension subtest of the Nelson Reading Skills Test. Several months earlier, she had obtained an IQ of 88 on the Otis-Lennon Mental Abilities Test. The appropriate steps set out above will now be followed, as needed for Harriet.

I. The Otis-Lennon is a general test of academic aptitude and as such it is not highly relevant to any one school subject. However, since this routine testing is being used only to form hypotheses about Harriet, its score may be used. Before reaching a firm decision about Harriet, however, a more relevant aptitude test should be used.

II-B.. The achievement percentile rank of 16 converts in Appendix A to a T-score of 40.

III-B. The Otis-Lennon manual provides grade-based percentile ranks. Harriet's is 31. This converts in Appendix A to a T-score of 45.

IV. 40 - 45 = -5.

V-A. The difference of -5 is rather minimal.

V-B. The Reading Comprehension subtest has a third-grade parallel-form reliability coefficient of .86 while that of the appropriate level of the Otis-Lennon is .90. (Measures of internal consistency are higher, but the parallel-form reliability estimates better reflect the several sources of error of measurement.) When Appendix D is entered with these values, the standard error of the difference is found to be 4.9 T-score units.

The difference of -5 is modest on both of the above bases. It may merit further testing, as will be discussed below, but it definitely does not warrant a firm conclusion.

Studying Intraindividual Differences in Achievement

The problems of judging the importance of differences in a student's achievement among subjects are similar to those considered above. In fact, if students' achievement profiles consisted of scores of only two tests, the tasks would be virtually identical. In comparing even two scores, errors of measurement can cause either score to overestimate or to underestimate the examinee's true status. This "double jeopardy" is the reason that the standard

error of a difference between two scores is always greater than the standard error of either separate score. If one measure is slightly overestimated while the other is somewhat underestimated, then a fairly large difference can result purely from measurement errors.

But profiles usually contain several scores. Suppose a student's profile contained scores in four subjects. If each score is compared with each other score, a total of 6 comparisons will be made. If another test were added to the battery, the number of comparisons would jump to 10. A serious problem arises when one looks for differences among scores in a profile; the chance factors have multiple opportunities to lead to faulty interpretations.

The more scores there are in a profile, the greater likelihood that errors of measurement will cause some differences to seem to be of practical and/or statistical significance when no real difference exists in the examinee's true achievement. This is what makes the "shotgun approach" or the "hunting expedition" in profile analysis likely to yield misinterpretations. If one gives several tests to almost any examinee, it is quite likely that at least one of the resulting differecnces will be big enough to attract attention.

There is obviously little merit in using procedures that identify most students as having disabilities. The intent is to narrow the group to those likely to benefit most from the instruction that a LD program can offer. There are two convenient ways in which the group can be reduced in size to eliminate persons who might otherwise be selected solely because of measurement errors.

First, a difference between attributes measured either by two achievement tests or by an achievement test and an aptitude test should be firmly concluded only when there is more than one source of evidence of its existence. Test should be used <u>either</u> to form hypotheses <u>or</u> to verify them, but not for

both at once. A student normally has been identified by screening; this involves rather specific hypotheses. Disability is customarily suspected in just one or two subjects, certainly not in all subjects. Achievement-aptitude differences should be obtained only in these one or two subjects. (There is not really anything inherently wrong, of course, with computing the discrepancies for other subjects. The point is that if they were computed in a blind "hunting expedition" and one difference were found to be great, then it should be treated as a hypothesis, not as an established finding.)

Second, in considering irregularities in an individual's achievement profile, the score for which a hypothesis of deviation is being investigated should be compared with the mean of several other scores in the profile, not with the most deviant or atypical of the other scores. These two precautions will go a long way toward minimizing the hazards of profile interpretation.

If a classroom teacher has reported that Pete performs adequately in everything except listening comprehension then (1) his achievement-aptitude difference should be considered only for listening comprehension and (2) his achievement profile may be examined to see if his listening comprehension score is markedly lower than the mean of his other scores, not to see if just any random pair of scores happen to differ significantly.

Jointly Considering Achievement-Aptitude and Achievement-Achievement Differences

Figure 1 provides a convenient way simultaneously to record and consider several achievement and aptitude scores of a student being considered for LD classification. It can, of course, also be used simply to compare a single achievement score with a relevant aptitude score. Before each score is graphed, it is expressed in T-score units based upon grade-based comparisons.

STUDENT'S NAME _____ BIRTHDATE _____

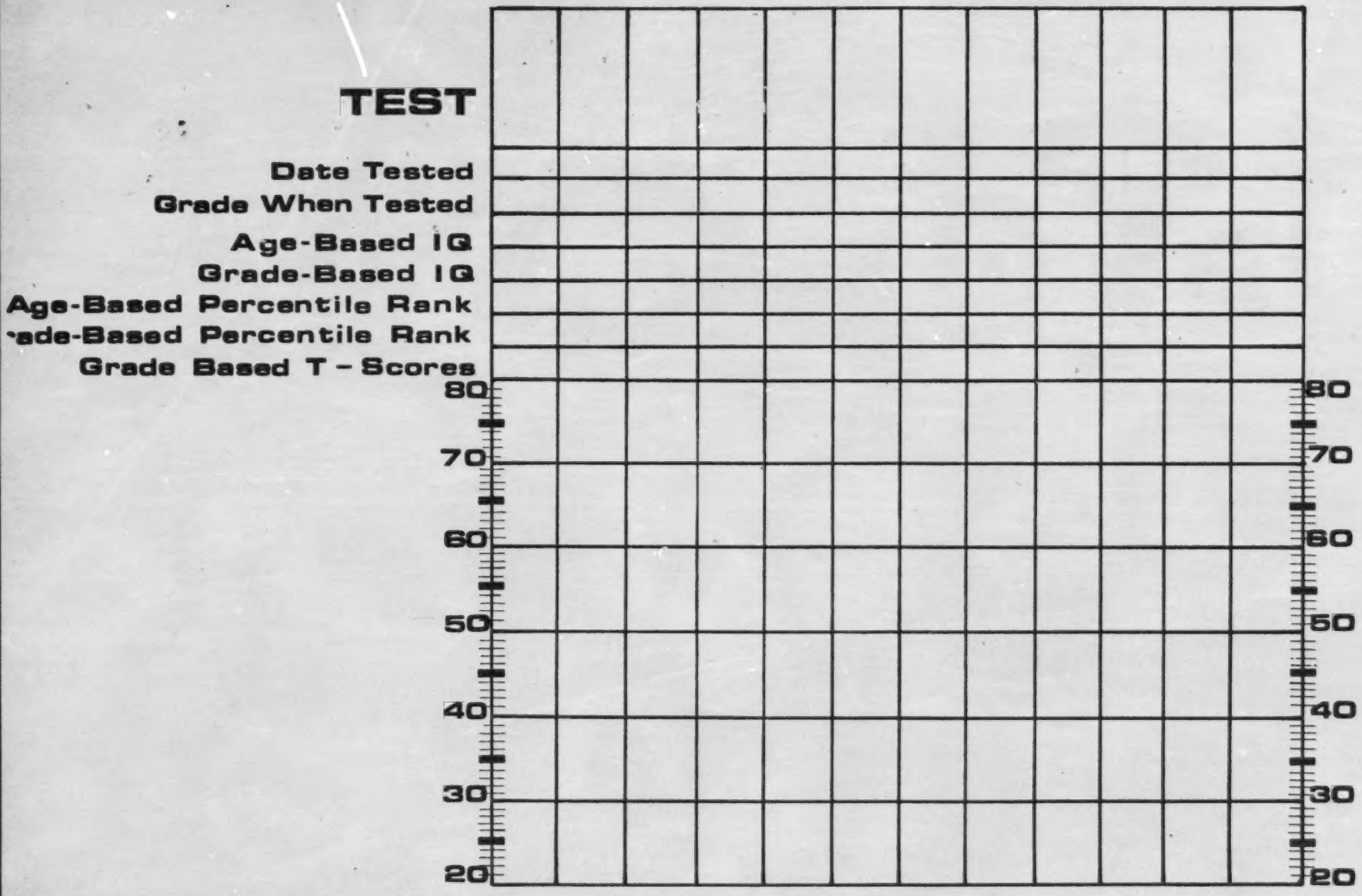

FIGURE 1 STUDENT PROFILE

- -

Insert Figure 1 about here.

- -

Returning to Harriet, suppose that after her modest achievement-aptitude discrepancy in reading comprehension was noticed, her teacher reported that she performs in class less well in reading than in any of her other subjects. As a result of referral, she was given the Peabody Individual Achievement Test near the beginning of fourth grade (4.1). Her grade-based percentile ranks were:

Mathematics	54
Reading Recognition	39
Reading Comprehension	14
Spelling	31
General Information	48

In addition, she was given the WISC-R at this time and her age-based deviation IQs were:

Verbal IQ	90
Performance IQ	90
Full Scale IQ	89

Harriet's results can now be recorded on the Student Profile Form as illustrated in Figure 2.* The conversions enable one to use the results of several tests that use different types of derived scores, some of which are

---

*If Harriet's WISC-R Verbal subtest raw scores for Information, Similarities, Arithmetic, Vocabulary and Comprehension are respectively 11, 11, 9, 34, and 12, they yield a Verbal Sum of Scaled Scores of 42 in the table (WISC-R Manual, p. 130) appropriate to her age, but a Verbal Sum of Scaled Scores of 48 in the table (WISC-R Manual p. 128) suitable for the median-aged student (9-7) whose grade placement is 4.1. These sums convert into the respective Verbal IQs of 90 and 97 shown in Figure 2.

age based and some of which are grade based. The scores have all been rendered comparable by use of the steps laid out above.

- -

Insert Figure 2 about here.

- -

It was noted earlier that Harriet's achievement-aptitude discrepancy in reading comprehension was -5. This rather minimally approached the first primary criterion for LD classification. If her PIAT Reading Comprehension T-score is compared with the T-score from her most relevant aptitude score, the Verbal IQ, then the difference is somewhat more pronounced, -9. This adds confidence to the earlier hypothesis of an achievement-aptitude discrepancy.

Note that if Harriet's age-based Verbal IQ of 90 (or T-score estimate of 43) had been compared with her PIAT Reading Comprehension percentile rank of 14 (or T-score equivalent of 39), her achievement-aptitude discrepancy would not have been revealed in its appropriate magnitude. This example illustrates the need to avoid making mismatched comparisons between grade-based and age-based scores.

Turning now to the teacher's report that Harriet's achievement was better in other subjects than in reading, the PIAT provides data with which to check this hypothesis. After the necessary conversions are entered on Figure 2, it is found that Harriet's PIAT Reading Comprehension T-score (39) and her mean on the other PIAT subtests (48) differ by 9 T-score units. She clearly meets the secondary criterion. Insofar as test data are concerned, she seems to be a good candidate for a LD program.

STUDENT'S NAME Harriet Fudd BIRTHDATE 9/17/69

TEST	Nelson RST Read. Comp.	Otis-Lennon	WISC-R Verbal	WISC-R Performance	WISC-R Full Scale	PIAT Math.	PIAT Read. Rec.	PIAT Read. Comp.	PIAT Spelling	PIAT Gen. Info.		
Date Tested	5/22/79	11/1/78	10/17/79	10/17/79	10/17/79	10/23/79	10/23/79	10/23/79	10/23/79	10/23/79		
Grade When Tested	3.8	3.2	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1		
Age-Based IQ		88	90	90	89							
Grade-Based IQ			97									
Age-Based Percentile Rank												
Grade-Based Percentile Rank	16	31				54	39	14	31	48		
Grade Based T-Scores	40	45	48			51	47	39	45	49		

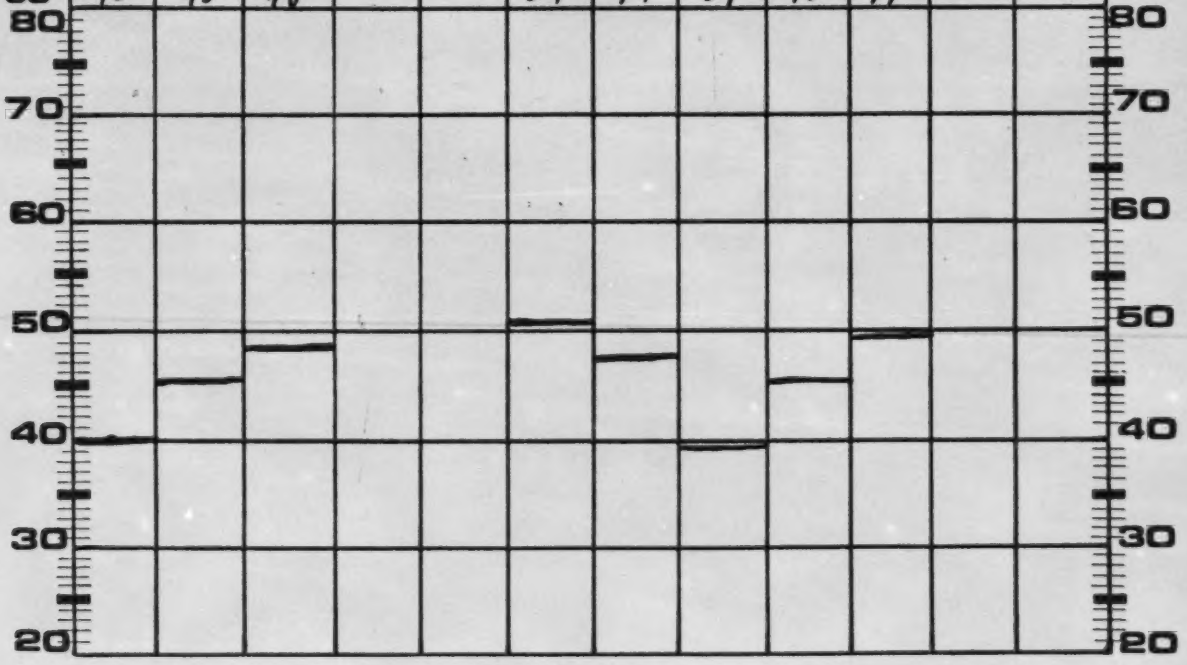

FIGURE 2-HARRIET'S PROFILE

Summary

A critical review of the literature dealing with the quantification of achievement-aptitude differences revealed that methods developed to date suffer from grave inadequacies. From the experience of past methods, principles were developed that (1) openly recognized the painful necessity to assume that the norms for well standardized tests are comparable, (2) stressed that aptitude tests used for LD classification should be relevant to their use, (3) recommended the use of T-scores for comparing results obtained from different tests, and (4) avoided the frequent problem of mixing age-based and grade-based normative data. Finally, a set of step-by-step procedures for comparing aptitude with achievement scores was offered, rationale for considering an individual's achievement profile were discussed, necessary tables were provided, and a figure was given with which a student's aptitude and achievement scores could be recorded, converted as needed, and graphed.

While many subjective factors should be considered when making LD placement decisions, at least one aspect can be subjected to reasonably objective analysis. It is hoped that the procedures presented in this article will enable better assessment of the achievement-aptitude discrepancies that PL 94-142 requires.

References

Bond, Guy L. & Tinker, Miles A. Reading Difficulties: Their Diagnosis and Corrections, Third Edition. New York: Appleton-Century-Crofts, 1973.

Cronbach, Lee J. Essentials of Psycholgical Testing, Third Edition. New York: Harper & Row, Publisher, 1970.

Erickson, Marilyn T. "The Z-Score Discrepancy Method for Identifying Reading Disabled Children." Journal of Learning Disabilities, 1975, 8, 308-312.

Harris, Albert J. How to Increase Reading Ability, Fourth Edition. New York: David McKay Company, Inc., 1961.

Harris, Albert J. & Sipay, Edward R. How to Increase Reading Ability, Sixth Edition. New York: David McKay Company, Inc., 1975.

Monroe, Marion. Children Who Cannot Read. Chicago: The University of Chicago Press, 1932.

Myklebust, Helmer R. Process in Learning Disabilities, Vol. 1. New York: Grune & Stratton, 1968.

Thorndike, Robert L. The Concepts of Over- and Underachievement. New York: Teachers College, Columbia Unversity, 1963.

Winkley, Carol K. "Building Staff Competence in Identifying Underachievers." In Robinson, Alan (Ed.) The Underachiever in Reading. Chicago: University of Chicago, Supplementary Educational Monograph No. 92, 1962.

Woodbury, Charles A. "The Identification of Underachieving Readers," Reading Teacher, 1963, 16, 218-223.

APPENDIX A

Percentile Ranks to T-Scores

Percentile Rank	T-Score	Percentile Rank	T-Score	Percentile Rank	T-Score
99+	76	66	54	32	45
99	73	65	54	31	45
98	71	64	54	30	45
97	69	63	53	29	44
96	68	62	53	28	44
95	66	61	53	27	44
94	66	60	53	26	44
93	65	59	52	25	43
92	64	58	52	24	43
91	63	57	52	23	43
90	63	56	52	22	42
89	62	55	51	21	42
88	62	54	51	20	42
87	61	53	51	19	41
86	61	52	51	18	41
85	60	51	50	17	40
84	60	50	50	16	40
83	60	49	50	15	40
82	59	48	49	14	39
81	59	47	49	13	39
80	58	46	49	12	38
79	58	45	49	11	38
78	58	44	48	10	37
77	57	43	48	9	37
76	57	42	48	8	36
75	57	41	48	7	35
74	56	40	47	6	34
73	56	39	47	5	33
72	56	38	47	4	32
71	56	37	47	3	31
70	55	36	46	2	29
69	55	35	46	1	27
68	55	34	46	1-	24
67	54	33	46		

APPENDIX B

Median Age by Grade*

Grade	Median Age	Grade	Median Age	Grade	Median Age
K.0	5-4	4.4	9-10	8.8	14-2
K.1	5-5	4.5	9-11	8.9	14-3
K.2	5-6	4.6	10-0	9.0	14-6
K.3	5-7	4.7	10-1	9.1	14-7
K.4	5-9	4.8	10-2	9.2	14-8
K.5	5-10	4.9	10-3	9.3	14-9
K.6	5-11	5.0	10-6	9.4	14-10
K.7	6-0	5.1	10-7	9.5	14-11
K.8	6-1	5.2	10-8	9.6	15-0
K.9	6-2	5.3	10-9	9.7	15-1
1.0	6-5	5.4	10-10	9.8	15-2
1.1	6-6	5.5	10-11	9.9	15-3
1.2	6-7	5.6	11-0	10.0	15.6
1.3	6-8	5.7	11-1	10.1	15.7
1.4	6-9	5.8	11-2	10.2	15-8
1.5	6-10	5.9	11-3	10.3	15-9
1.6	6-11	6.0	11-6	10.4	15-10
1.7	7-0	6.1	11-7	10.5	15-11
1.8	7-1	6.2	11-8	10.6	16-0
1.9	7-2	6.3	11-9	10.7	16-1
2.0	7-5	6.4	11-10	10 8	16-2
2.1	7-6	6.5	11-11	10.9	16-3
2.2	7-7	6.6	12-0	11.0	16-6
2.3	7-8	6.7	12-1	11.1	16-7
2.4	7-9	6.8	12-2	11.2	16-8
2.5	7-10	6.9	12-3	11.3	16-9
2.6	7-11	7.0	12-6	11.4	16-10
2.7	8-0	7.1	12-7	11.5	16-11
2.8	8-2	7.2	12-8	11.6	17-0
2.9	8-3	7.3	12-9	11.7	17-1
3.0	8-6	7.4	12-10	11.8	17-2
3.1	8-7	7.5	12-11	11.9	17-3
3.2	8-8	7.6	13-0	12.0	17-6
3.3	8-9	7.7	13-1	12.1	17-7
3.4	8-10	7.8	13-2	12.2	17-8
3.5	8-11	7.9	13-3	12.3	17-9
3.6	9-0	8.0	13-6	12.4	17-10
3.7	9-1	8.1	13-7	12.5	17-11
3.8	9-2	8.2	13-8	12.6	18-0
3.9	9-3	8.3	13-9	12.7	18-1
4.0	9-6	8.4	13-10	12.8	18-2
4.1	9-7	8.5	13-11	12.9	18-3
4.2	9-8	8.6	14-0		
4.3	9-9	8.7	14-1		

*Interpolated and extrapolated from U.S. Bureau of the Census, Current Population Reports, Series P-20, No. 319, "School Enrollment-Social and Economic Characteristics of Students: October 1976," U.S. Government Printing Office, Washington, D.C., 1978.

APPENDIX C

Deviation IQs to T-Scores

DIQ SD16	DIQ SD15	T-Score	DIQ SD16	DIQ SD15	T-Score
148	145	80	98-99	98-99	49
146-147	143-144	79	96-97	97	48
144-145	142	78	95	95-96	47
143	140-141	77	93-94	94	46
141-142	139	76	92	92-93	45
140	137-138	75	90-91	91	44
138-139	136	74	88-89	89-90	43
136-137	134-135	73	87	88	42
135	133	72	85-86	86-87	41
133-134	131-132	71	84	85	40
132	130	70	82-83	83-84	39
130-131	128-129	69	80-81	82	38
128-129	127	68	79	80-81	37
127	125-126	67	77-78	79	36
125-126	124	66	76	77-78	35
124	122-123	65	74-75	76	34
122-123	121	64	72-73	74-75	33
120-121	119-120	63	71	73	32
119	118	62	69-70	71-72	31
117-118	116-117	61	68	70	30
116	115	60	66-67	68-69	29
114-115	113-114	59	64-65	67	28
112-113	112	58	63	65-66	27
111	110-111	57	61-62	64	26
109-110	109	56	60	62-63	25
108	107-108	55	58-59	61	24
106-107	106	54	56-57	59-60	23
104-105	104-105	53	55	58	22
103	103	52	53-54	56-57	21
101-102	101-102	51	52	55	20
100	100	50			

APPENDIX D

Standard Errors of Differences in T-Score Units

Reliability of Other Tests	Reliability of One Test														
	.98	.96	.94	.92	.90	.88	.86	.84	.82	.80	.78	.76	.74	.72	.70
.98	2.0	2.5	2.8	3.2	3.5	3.7	4.0	4.2	4.5	4.7	4.9	5.1	5.3	5.5	5.7
.96	2.5	2.8	3.2	3.5	3.7	4.0	4.2	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.8
.94	2.8	3.2	3.5	3.7	4.0	4.2	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.8	6.0
.92	3.2	3.5	3.7	4.0	4.2	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.8	6.0	6.2
.90	3.5	3.7	4.0	4.2	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.8	6.0	6.2	6.3
.88	3.7	4.0	4.2	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.8	6.0	6.2	6.3	6.5
.86	4.0	4.2	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.8	6.0	6.2	6.3	6.5	6.6
.84	4.2	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.8	6.0	6.2	6.3	6.5	6.6	6.8
.82	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.8	6.0	6.2	6.3	6.5	6.6	6.8	6.9
.80	4.7	4.9	5.1	5.3	5.5	5.7	5.8	6.0	6.2	6.3	6.5	6.6	6.8	6.9	7.1
.78	4.9	5.1	5.3	5.5	5.7	5.8	6.0	6.2	6.3	6.5	6.6	6.8	6.9	7.1	7.2
.76	5.1	5.3	5.5	5.7	5.8	6.0	6.2	6.3	6.5	6.6	6.8	6.9	7.1	7.2	7.3
.74	5.3	5.5	5.7	5.8	6.0	6.2	6.3	6.5	6.6	6.8	6.9	7.1	7.2	7.3	7.5
.72	5.5	5.7	5.8	6.0	6.2	6.3	6.5	6.6	6.8	6.9	7.1	7.2	7.3	7.5	7.6
.70	5.7	5.8	6.0	6.2	6.3	6.5	6.6	6.8	6.9	7.1	7.2	7.3	7.5	7.6	7.7